TOWARDS A NATURAL MORPHOLOGY OF COMPOUNDING

1. Introduction

Although compounding is the part of morphology which is closest to syntax, studies in Natural Morphology and Natural Syntax, as pioneered by Mayerthaler (1981) and Orešnik (2004) have rather avoided the field of composition. This contribution represents a modest tentative towards remedying this lacuna.

The theory of Natural Morphology (NM, cf. Kilani-Schoch & Dressler 2005, Dressler et al. 1987, Dressler 2000) takes naturalness as a cover term for a set of more specific terms which are defined in three subtheories: 1. a universal markedness theory of system-independent morphological naturalness (cf. Mayerthaler 1981), focussing on universal preferences, 2. a theory of typological adequacy (cf. Dressler 1985, 1988a), 3. a theory of system-dependent naturalness or system-adequacy (cf. Wurzel 1984, Dressler & Ladányi 2000). This contribution represents a tentative approach towards the study of compounding limited to the first subtheory. Due to lack of space it will be highly selective. Therefore if a phenomenon can be well illustrated with an English, only English examples will be given.

This subtheory of universal markedness is a preference theory (cf. Vennemann 1983, Dressler 1999), which does not refer to any global or overall preference, but to what is universally preferred separately on each of a restricted number of naturalness parameters established deductively by universal markedness theory. Each of the following parameters and its preference degrees are deduced from extralinguistic bases via the metatheory of Peircean semiotics.

2. iconicity

The best-known semiotically-based parameter is the parameter of iconicity, based on Peirce's (1965) concept of icons with its most important hypoicons, i.e. diagrams. Diagrams represent an analogy between the relations of the signans and the signatum. Thus the fact that in the compound corner stone the second element stone is both the morphosemantic and the morphotactic head, whereas the first element corner is the non-head on both levels (i.e. signans and signatum, and the same holds, with reversed roles, for its twin compound stone corner) is diagrammatic. For it holds that a corner stone is a type of stone, with which it shares its syntactic and semantic features of being an inanimate, concrete noun (morphosemantic headhood) and that the plural suffix is affixed to the same second element (morphotactic headhood), and this establishes an exact, diagrammatic parallel between the morphotactic and the morphosemantic head-nonhead relations in the signatum and the sig-
nans. Such optimal diagrammaticity holds, for example, for most English compounds and for all the productive ones, but not for the unproductive type pickpocket, daredevil, where in the signatum the first element governs the second one, whereas the plural suffix is still added to the second element.

3. Indexicality

Another parameter derived from Peircean semiotics is the parameter of indexicality. An index is a sign where the signans directly refers to the signatum. All of morphology is indexical insofar a morphological marker refers to the base of the rule that introduces it. This fulfills the semiotic definition of an index which refers deictically to its nearby object. A specific instance of indexicality are linking elements or interfixes (see § 5).

Most of the compounds in all languages have their head within the compound, thus they are endocentric compounds, such as in E. black-board, which is a type of board. Exocentric compounds have their head outsides or more precisely, the head has to be inferred. Thus a loud-mouth is a person who, metaphorically, has a loud mouth, bare-foot an adjective derived from a noun-phrase, pick-pocket is a person who so-to-say picks pockets, and a pass-port was originally a document which allows to pass a port. Endocentric compounds are by far preferred to exocentric compounds in the languages of the world, because they allow much easier access to the head, i.e. they are more natural on the parameter of indexicality. Sometimes endocentric and exocentric compounds are morphologically differentiated. For example, the English plural sabre teeth refers to teeth (endocentric), the plural sabre tooth-s (exocentric) to animals having sabre teeth.

4. Morphosemantic transparency

From the semiotic preference for transparency (Koj 1979), we can derive the two parameters of morphosemantic and (§ 5) morphotactic transparency. On the parameter of morphosemantic transparency, full transparency means fully compositional meaning, as is generally the case with inflectional meanings. Motivation of a complex or derived word by its parts or by applying a word-formation rule to its bases (or bases in case of compounds) is best served in case of optimal transparency.

In word formation, morphosemantic transparency can never be complete, because Frege's principle of semantic compositionality can hold only for syntax where the meaning of a syntactic unit is typically fully derivable from the meanings of its syntactic constituents (except in idiomatic phrases, such as to kick the bucket). This does not hold for word formation, insofar as all accepted words are stored and thus lexicalised (Bauer 1983, Meyer 1992), whereas not yet accepted neologisms, generally, realise only one of the potential meanings of a compound or derivation. Thus we must differentiate between transparent word formation meaning (G. Wortbildungsbedeutung, cf. Corbin's 1987 notion of sens construit) and lexicalised word
meaning. Hence word formation rules can only predict word formation meaning but not the opacifying differences between word meaning and word formation meaning.

The end point of opacification (= non-transparency) is fossilization, where members and their combination are hardly visible, if at all. For example, E. lord and lady are not recognizable as compounds any more, E. nostril, gospel, handsel barely for the first member, if at all. In contrast, E. dandelion may be marginally recognized as a (non-prototypical) compound, whose first member is difficult to identify, whereas its second is semantically totally opaque (diachronically a loan from the French idiomatic phrase *dent de lion*, lit. tooth of lion'). In contrast, the German interfixed equivalent Löw-en-zahn is a prototypical compound (cf. § 8). Less lexicalization means more transparency, more lexicalisation more opacity. More transparency implies more motivation of the compound via its members.

Due to Frege's principle of compositionality, the meaning of a non-idiomatic noun-phrase, such as a high school is fully compositional and thus transparent, the cognate compound high-school is not. Still the compound high-school may be classified as a morphosemantically transparent compound, because the meaning of the head is fully transparent and because the semantic motivation by its first member (non-head) is still evident: a high-school is high in a metaphorical sense, i.e. relatively high in contrast to elementary and grade school. In fact, a high-school is a specific instance of what a *high school* may mean potentially. Thus, in a first approximation, we may define an actual transparent compound as one whose meaning is a subset of the set of potential meanings of the compound as constructed grammatically via the combination of the meanings of the two compound members. This is more precise than Shaw's (1979) criterion that the head must be a hyperonym of the compound. This is also relevant for psycholinguistic research on non-existing, but potential compounds. They are always transparent, but their use depends on the possibility to instantiate via (e.g. metaphorical) inferences a pragmatically plausible potential meaning.

Based on this approach and following Libben (1998) we can differentiate the following four fundamental degrees of morphosemantic transparency (which are clearer and more systematic than Shaw's 1979):

1) transparency of both members of the compound, e.g. *door-bell*,
2) transparency of the head member, opacity of the non-head member, e.g. *straw-berry*,
3) transparency of the non-head member, opacity of the head member, e.g. *jail-bird*,
4) opacity of both members of the compound, e.g. *hum- bug*.

This scale of transparency presupposes that transparency of the head is more important than of the non-head.

Further refinement is possible, if we differentiate between direct and indirect, i.e. metaphoric motivation (cf. de Knop 1987), which would allow to divide degrees 1) – 4) above into sub-degrees with metaphorically motivated vs. unmotivated opacity. Thus ratchet tooth has a transparent first member, but a metaphorically motivated second member (the leaves are compared with teeth), thus belonging to the less transparent variant of transparency degree 1).
In principle, compounds are morphosemantically more transparent than derivations, because they are more descriptive (cf. Seiler 1991), for example, the compound *dishwashing machine* tells the language user more about what it is, than the derivation *dishwash-er*, i.e. the compound serves better the function of semantic motivation than the corresponding derivation, cf. Crocco-Galeas & Dressler (1992) for the Italian preference for coining compounds as terms for low-technology machines vs. derivations for high-technology machines. Therefore, in the process of first language acquisition, several children have been observed to replace temporarily the less transparent (and ambiguous) agent and instrument noun *open-er* with *open-man* and *open-thing* respectively (Clark, Hecht & Mulford 1986), cf. It. *post-ino* ‘postman’ ➔ *porta-posta* ‘lit. post-carrier’ (Lo Duca 1990).

5. Morphotactic transparency

On the parameter of morphotactic transparency, the most natural forms are those where there is no opacifying obstruction to ease of perception.

Bases of word-formation rules are, ceteris paribus, morphotactically and morphosemantically most transparent, if they are autonomous words in their uninflected form, already less so in their inflected citation form (Dressler 1988b, cf. Rainer 1993: 98ff). This universal preference for word-based morphology applies to compounding even more than to inflection and derivation.

Larger and thus also universally less preferred bases are represented by inflected words which are not identical with citation forms. Such bases are very rare and restricted in English, e.g. *sport-s-man, sale-s tax* with a pluralised first base (cf. Jensen 1990). Pluralised second bases are wide-spread in Romance languages.

Still larger bases, namely phrases appear in synthetic compounds of the types E. *dish-wash-er* and *lion-heart-ed, good-natur-ed* (cf. Crocco Galéas 2003) in the type E. *three-star general, three-phase motor, three-color process*, etc., where the modifier of the noun-phrase in the non-head position is often difficult to omit (e.g. E. *star general* belongs to a different compound type and designates a „general who is a star“). For other phrasal-compound patterns cf. Lieber (1992: 92ff) and Booij (1992: 45ff). Whole sentences as first members of a binary compound occur only rarely, such as in E. *do-it-yourself movement*, which is less natural than *eye movement*. Complex sentences occur only in consciously formed occasionalisms, such as *an oh-what-a-wicked-world-this-is-and-how-I-wish-I-could-do-something-to-make-it-better-and-nobler-expression* (J.K. Jerome: Three Men in a Boat).

Also smaller bases than words are dispreferred. The smallest possible bases are roots. Examples of root-based compounding in English are neoclassical compounds, such as *bio-chemistry, bio-acoustical*. They have a wider occurrence in other languages, e.g. in the German unproductive pattern *Fried-hof* †cemetery (lit. peace-court)† or *Kirsch-baum* †cherry-tree†, where the citation form of the first base is *Friede(n), Kirsche*. The alternative of assuming phonological deletion of the final vowel is less attractive in view of the compounds *Fried-en-s-vertrag* †peace treaty†, *Kirsch-en-baum*
between morphosemantic and morphotactic transparency/opacity. For example, comparable nominal compounds tend to be morphotactically opaque when they are morphosemantically opaque. For example, the first base is morphosemantically more transparent in E. *mother-land* and *main-land* than in *Dixie-land*. In analogical parallelism in morphotactics, the phonological shape of the second base has more rarely a reduced (and thus morphotactically more opaque) vowel in the first and second than in the last compound. Or the morphotactically transparent Breton compound *ti-forn* is also morphosemantically transparent: ‘house (with an) oven’. Its variant *ti-vorn* ‘bakery’ (with lenition /f/ → /v/) is both morphosemantically and morphotactically less transparent. The psychological reality of this difference becomes obvious in different plural formation: Pl. *ti-erforn*, where the plural suffix is attached to the transparent lefthand head, vs. Pl. *ti-vorn-iou*, where the plural suffix is attached word-finally, because this doubly opacified word is less easy to decompose.


6. Biuniqueness

Another semiotically-based parameter has biuniqueness as its most natural option. Biuniqueness, which holds if one and the same form has always the same meaning (and vice-versa), is more natural than uniqueness and especially ambiguity.
Biuniqueness can be obtained more easily in terminology (Felber & Budin 1989: 122f). However even here terminologists do not strive for obtaining general biuniqueness but only terms which are biunique within each scientific domain (in the horizontal stratification of languages for specialized purposes), particularly for the most abstract level of theoretical argumentation (in the vertical stratification of LSP). Thus it is unproblematic that a term such as *morphology* has a very different meaning in linguistics and in medicine or biology, as long as it has always the same meaning in linguistics and if this linguistic concept is not expressed by another term as well. The German corresponding synonyms *Morphologie* and *Formenlehre* are only an apparent counter-example, because the second term is typical for popular science, i.e. for a lower technolectal level of abstractness. Sciences where different terms are indiscriminately used for the same concept (as often in pedagogy) are likely to get a low rating for such usage. Since, however, different schools of a discipline often tend to use the same terms differently, the quest for biuniqueness must be limited to the same text world, i.e. biuniqueness of a term should hold at least within the same text (cf. Dressler 1994: 956f). In addition we find a tendency to enhance biuniqueness of the morphologically decomposed parts of words, such as compound members, e.g. in attributing the sense of an applicative descriptive science to the combining form *-graphy*, as in *lexico-graphy, termino-graphy* vs. *lexicology, terminology*.

7. Figure-ground preferences

Another universal preference is the tendency towards figure-ground sharpening (Scherer 1984), i.e. for contrasting syntagmatically between a more important, more dynamic and clearer foreground and a less important, more static and fuzzier background. In word formation the head represents the figure, the non-head the ground. Therefore, as we have seen in § 4, morphosemantic transparency of the head is more important than that of the non-head. This is also another reason why endocentric compounds are preferred to exocentric ones (§ 3).

The figure-ground distinction between head and non-head is established by morphosemantic and morphotactic subordination of a non-head under the head (cf. § 2). This is by far the preferred situation in suffixation and compounding. The preference for subordination of a non-head to a head holds even in exocentric compounds. Not only is the non-head morphosemantically subordinated to the absent head. But subordination holds, secondarily, even among the actually present members of exocentric compounds: *loud* is subordinated to *mouth* in *loud-mouth*, whereas in *pick-pocket* we find again non-uniformity of (secondary) head-hood: although the primary, semantic head which designates the person who picks pockets, is not expressed, *pocket* is, secondarily, the morphotactic head which determines inflection, whereas syntactically *pocket* is subordinated to *pick*.

In compounding, subordinate compounds are universally preferred over coordinate compounds which have two or more morphosemantic heads (cf. Wunder-
lich 1986: 241), i.e. without the clear figure-ground distinction of subordinate compounds. Thus in E. speaker-hearer both members are of equal status, although the plural ending attaches only to the right member, another example of non-uniform headhood. Coordinate (or coordinative) compounds may again be endocentric, such as speaker-hearer or the adjective bitter-sweet, also called appositional compounds, or they may be exocentric, such as morphology-syntax interface, where the two coordinated compound members have their semantic head outsides: it coincides with interface, the syntactic head of the whole noun-phrase. More subtle properties of coordinate compounds may differ considerably from language to language (cf. Olsen 2001): The linear order of members in coordinate compounds is not grammatically determined (since all members are equipollent), but pragmatically (e.g. the most important first) or stylistically, e.g. prosodically (e.g. the longest last). The first reason explains the order of speaker-hearer (because linguists tend to think more of the speaker than of the hearer, cf. the term native speaker), the second explains why the order of the synonym speaker-listener is even more difficult to reverse (?listener-speaker) than in the case of speaker-hearer.

On the syntagmatic level, this lexical-pragmatical grading of importance in speaker-hearer is antagonistic to the inverse morphological order of non-heads followed by heads in English and the majority of languages. The latter sequence represents a universal preference for heads to be on the right side of non-heads, called the righthand head rule by Williams (1981: 248). A minority of languages has large classes of left-hand-headed compounds as well (cf. Zwanenburg 1992a, b, Scalise 1992: 179ff, Rai- ner 1993: 57), e.g. It. capo-stazione 'station-master' (lit. 'head station'). This recalls the suffixing preference, whereby suffixes are preferred to prefixes. Most explanations of the suffixing preference start from the assumption that it is better for a word to start with the lexical basis (cf. Hall 1992), which would hold for both right-headed and left-headed compounds. Thus if one compares compounds, then the more valid generalisation seems to be that it is better for complex words to end with the head. This would also explain the tendency for prefixes not to be heads (cf. Hall 1992). An explanation for the right-hand head preference may lie in the psycholinguistic recency effect which makes the end of a word more salient, which is especially important in early language acquisition (cf. Wijnen et al. 2001).

8. Binarity

Semiotic and thus also grammatical relations are preferentially binary, as already observed by Peirce (1965: II.277), a preference which is maybe based on the binary nature of neurological information transmittance), and this both in paradigmatics and syntagmatics. As has been ascertained already within structural linguistics, even apparently ternary relations have to be usually split up into two binary relations. For example, morphology has to be subdivided into inflectional morphology and word formation, the latter in its turn into derivation and compounding.

In syntagmatic relations, the preferred patterning consists in concatenating one element to one base, as in affixation, or by compounding two bases, both in subor-
dinate compounds and coordinate compounding. In subordinate compounding there is a statistical preference for two-member compounds over more-member compounds, but in their derivational history are nearly always formed via concatenating one new element or two-member compound to another element or two-member compound, as in the English three-member compound three-star-general, which has to be grouped into two binary relations, i.e. [[[three][star]][general]], where the phrase [three][star] forms an intermediate unit. This may be done recursively, as in G. Donau-dampf-schiff-fahrt-s-gesellschaft ‘Danube steam shipping society’ as [[Donau][[Dampf][Schiff][Fahrt]]][Gesellschaft]].

This preference is true as well for coordinated compounds, which therefore have been called Dvandva ‘pair’ by the ancient Indian grammarian Pāṇini. This holds for E. queen-mother or prince-consort, but (due to entirely extralinguistic reasons) not for flags, e.g. red-white-red for the Austrian and Peruvian flag.

9. Prototypes

The theory of NM also uses the notion of prototypes (cf. Dressler 1990). A prototypical compound is a word which consists of two prototypical lexical words. This excludes combinations of clitics, such as It. glielo = gli ,to him' + lo ,it', because clitics are non-prototypical words or, according to Anderson (1992) phrasal affixes. As a consequence, also combinations of a lexical word and a clitic are excluded, e.g. It. darlo ,to give it'. Prototypically all members of a compound recur as free forms (i.e. autonomous words). This is not the case with so-called cranberry morphs (cf. Aronoff 1976: 10). The first members of E. cran-berry, boysen-berry, huckle-berry do not occur independently, but they can be assigned a lexical meaning which distinguishes them from those of other berries. Also neoclassical compounds (cf. Bauer 1998) may contain non-autonomous parts, so-called combining forms, e.g. helio-gra-phy, biblio-therapy, biblio-graphy (cf. Iacobini 1998). Prototypically, compound members belong to major lexical categories, with a preference for nouns. Combinations of function words, such as E. wherever, with-in, on-to, and with function words, such as G. Ich-sucht ‘ego-mania’ are rare and unproductive.

The core of prototypical compounds is surrounded by non-prototypical compounds in transition to syntax, derivation, and the lexicon. Most important representatives of the first transition area are so-called juxtapositions which share with compounds only the syntactic criterion of non-separability, i.e. by the impossibility of changing linear order and of inserting another word between the members of a compound vs. of a syntactic phrase. But phonologically and morphologically, juxtapositions may have properties of syntactic phrases. For example, Ge. der Hohe-priester, the (Jewish) high-priest' with main stress on the first member (modifier) and secondary stress on the second (head) may inflect either only the final part, as in Gen. des Hohe-priester-s or both members, i.e. des Hohe-n-priester-s, without any meaning difference. The second variant is more phrase-like, the first more compound-like. The corresponding prototypical compound would be (potential) Hoch-priester. Compare the single or double inflection of the Polish toponym Bialy-stok

36
(lit. 'white slope'), Gen. Bialy-stok-u or Bialego-stok-u (cf. adj. masc. bialy, Gen. bial-ego 'white'). A partially similar difference exists between more phrase-like Italian "loose compounds" vs. prototypical "strict compounds" (Scalise 1992: 180-181), e.g. mezza-luna 'half-moon', Pl. mezze-lune vs. lung-arno, Pl. lung-arn-i, lungo-tevere, Pl. lungo-tever-i. So far for the gradual transition between prototypical syntactic phrases and prototypical compounds.

Juxtapositions may be understood as diachronic transition stages in the grammaticalization of syntactic phrases to compounds. Also within the further grammaticalization process of phrases into derivatives, whereby one part of the phrase changes into an affix, we find a transition stage, the so-called semi-affixes or affixoids (cf. Wolff 1984: 89ff). In all these cases of varying degrees of grammaticalisation, the more grammaticalised a construction is, the less it tends to be (morpho)semantically and formally (e.g. morphotactically) transparent (cf. Lehmann 1995).

Well-known instances are semi-suffixes/suffixoids as representatives of the transition between compounds and suffixation: the German productive suffixoid -tum, cognate of E. doom and the suffix -dom, derived from it, was originally identical with an independent, now obsolete noun. It is a prima-facie suffix, but with the two following characteristics of compound heads: 1) it carries a secondary stress, 2) it is preceded by the compound-linking (interfix) -en- (cf. Fuhrhop 1998 and above § 5). The same holds for the suffixoid -schaft (cognate of E. -ship), e.g. in Frau-en-schaft 'women's organization' (lit. 'women-ship'), Frau-en-tum 'women-dom', cf. the compounds Frau-en-arzt 'women's doctor, gynecologist', Frau-en-feind 'woman hater' vs. the derivatives Frau-chen 'little woman', frau-lich 'womanly'.

In the transition area between compounding and the lexicon we find words which are only marginally identifiable as compounds, such as E. dandelion (cf. § 4).

10. Conclusion

Since the space available has been very limited, the above treatment of universal compound preferences had to be rather sketchy. But it should have become clear that these preferences, which hold for the whole area of morphology, are deducted from higher-order principles and can account for actual distribution preferences. As is well-known, compounds show many parallels to syntax (cf. Borer 1988). But this leads into the research domain of Janez Orešnik to whom I dedicate this study.
Povzetek

PRISPEVEK K NARAVNEMU OBLIKOSLOVJU ZLOŽENK

Zloženke so v naravnem oblikoslovju še neobdelano področje. Avtor jih označi z vidika ikoničnosti, indeksikalnosti, morfosemantične in morfosintaktične prosojnosti, enoumnosti, razlikovanja med ospredjem in ozadjem, binarnosti, prototipičnosti. V tem okviru dožene vrsto univerzalnih teženj, o katerih pa sodi, da veljajo za vse oblikoslovje. Domneva, da bi se o zloženkah dale izreči dodatne univerzalne težnje, ko bi zloženke pogledali iz zornega kota skladnje.